Zhiyong Lu, Ph.D.

CONTACT INFORMATION	DHHS/NIH/NLM/NCBI 8600 Rockville Pike Building 38A, Room 1003A Bethesda, MD 20894	*Tel:* (301) 594-7089 *Fax:* (301) 480-2288 *E-mail:* luzh@ncbi.nlm.nih.gov http://www.ncbi.nlm.nih.gov/CBBresearch/Lu

RESEARCH INTERESTS
Bioinformatics, Biomedical Informatics
Text Mining (Named Entity Recognition; Biomedical Information Retrieval)
Machine Learning, Data Mining

EMPLOYMENT
Staff Scientist, National Center for Biotechnology Information (NCBI), 2007 -

EDUCATION
Ph.D. in Bioinformatics, University of Colorado Health Sciences Center, 2007
PhD Thesis: Text Mining on GeneRIFs, advised by Lawrence Hunter

M.S. in Computer Science, University of Alberta, 2003
MS Thesis: Predicting Protein Sub-cellular Localization from Homologs Using Machine Learning Algorithms, supervised by Duane Szafron and Russell Greiner

B.S. in Computer Science, Nanjing University, 2001

HONORS AND AWARDS
Individual Special Service Award, for excellence in performance of PubMed log analyses and development of new and valuable user services, National Library of Medicine, 2009

AMIA Annual Symposium 2008 Outstanding Paper Submission, 2008

Best performance in automatic protein-protein interaction extraction task, BioCreative II global competition, 2007

Conference (PSB 2007) Travel Award, Int'l Society of Computational Biology, 2007

Conference (ISMB 2006) Travel Award, Int'l Society of Computational Biology, 2006

Conference (PSB 2006) Travel Award, Int'l Society of Computational Biology, 2006

Conference (ISMB 2005) Travel Award, Int'l Society of Computational Biology, 2005

Professional Development Award, University of Colorado Health Sciences Center Student Senete, 2005

Conference (PSB 2005) Travel Award, Int'l Society of Computational Biology, 2005

Conference (Rocky 03) Travel Award, University of Colorado Center of Computational Biology, 2003

Undergraduate (merit-based) Student Scholarship, Nanjing University, 1997 - 2001

PROFESSIONAL SERVICES
Conference Organizing:
Co-organizer for the third BioCreative global challenge

Review and Referee Work for Journals:
Bioinformatics
BMC Bioinformatics

Journal of Biomedical Informatics
ACM Transactions on Information Systems
DATABASE - The Journal of Biological Databases and Curation

Scientific Program Committee for Conferences:
ACL-HLT BioNLP workshop, 2010
AMIA 2009 Annual Symposium, San Francisco USA, 2009
International Joint Conferences on Artificial Intelligence (IJCAI), Pasadena, CA, 2009
ISMB/ECCB 2009 Stockholm, Sweden, 2009
The Asia Pacific Bioinformatics Conference (APBC), 2009
NAACL-HLT BioNLP workshop, Boulder, CO 2009
ACL-HLT BioNLP workshop, Columbus, Ohio 2008
NAACL-HLT BioNLP workshop, Brooklyn, NY 2006
Pacific Symposium on Biocomputing (PSB), Maui, HI 2006

MENTORSHIP   **Co-supervised PhD Students:**
Craig Murray, University of Maryland at College Park, January 2008 - Present

Research Fellows:
Rezarta I. Dogan, Ph.D. August 2008 - present
Aurelie Neveol, Ph.D. October 2008 - present

Others:
Jiao Li, Visiting scholar from Department of Computer Science, Tsinghua University, August 2009 - present

TEACHING   Tutorial: Introduction to biomedical natural language processing, *The Asia Pacific Bioin-*
EXPERIENCE   *formatics Conference (APBC) 2009*, Beijing, China. January, 2009.

05/06-11/06 University of Colorado School of Medicine   Graduate Student Supervisor
Supervise summer intern students and rotation graduate students

08/04-11/04 University of Colorado School of Medicine   Teaching Assistant
BIOI 7711: Introduction to Bioinformatics (graduate level class)

09/02-12/02 University of Alberta   Teaching Assistant
CMPUT 366: Introduction to Artificial Intelligence

01/02-04/02 University of Alberta   Teaching Assistant
CMPUT 201: Practical Programming Methodology

09/01-12/01 University of Alberta   Teaching Assistant
CMPUT 101: Introduction to Computing

OTHER SERVICES Member of the Association of Computational Linguistics (ACL), 2008 -
AND   Member of the International Society of Computational Biology (ISCB), 2005 -
MEMBERSHIPS   Student Representative in Department Faculty Meeting, University of Colorado, 2004

Treasurer-elect, Computer Science Graduate Student Association, University of Alberta, 2003

Councilor-elect, Graduate Student Association, University of Alberta, 2003

Member of the American Association for Artificial Intelligence (AAAI), 2002

INVITED LECTURES

Text Mining for Omics, Genomics and Proteomics Interest Group, George Washington University, March 2009.

Evaluating IR Techniques for Enhancing MEDLINE Retrieval, Computational Bioscience Program, University of Colorado School of Medicine, December, 2008.

Evaluating Relevance Ranking Strategies for MEDLINE Retrieval, NLM Informatics Series, December, 2008

Evaluating and Enhancing NCBI Textual Resources, CBCB Seminar, University of Maryland College Park, November 2008

Evaluating and Enhancing NCBI Textual Resources, AI Seminar, Department of Computer Science, University of Alberta, April 2008

Text Mining on GeneRIFs, CLIP Colloquium, University of Maryland College Park, November 2007

Text Mining for Biology, Center for Bioinformatics, Peking University, June 2007

Text Mining for Biology, School of Life Sciences, Suzhou University, February 2007

Extracting Protein Transport Information from GeneRIFs, Bioinformatics supergroup, University of Colorado at Boulder, March 2006

Case Study: Protein Translocation, Biomedical Ontology Session Discussion Panel, Pacific Symposium on Biocomputing, January 2005

PEER REVIEWED PUBLICATIONS

1. Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, Chetvernin V, Church DM, Dicuccio M, Federhen S, Feolo M, Geer LY, Helmberg W, Kapustin Y, Landsman D, Lipman DJ, **Lu Z**, Madden TL, Madej T, Maglott DR, Marchler-Bauer A, Miller V, Mizrachi I, Ostell J, Panchenko A, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Shumway M, Sirotkin K, Slotta D, Souvorov A, Starchenko G, Tatusova TA, Wagner L, Wang Y, John Wilbur W, Yaschenko E, Ye J. Database resources of the National Center for Biotechnology Information, *Nucleic Acids Research*, Nov 2009

2. Dogan RI, Murray GC, Neveol A, **Lu Z***, Understanding PubMed user search behavior through log analysis, *to appear in Database*, 2009

3. **Lu Z***, Xie N, Wilbur WJ: Identifying related journals through log analysis, *Bioinfor-*

---

*Corresponding author

matics, 25(22):3038-9, 2009. (currently used in NLM's Journals database)

4. **Lu Z***, Wilbur WJ, McEntyre JR, Iskhakov A, Szilagyi L: Finding query suggestions for PubMed, to appear in *proceedings of the AMIA Annual Symposium 2009*. (currently used in PubMed)

5. Jin F, Huang M, **Lu Z**, Zhu X: Towards Automatic Generation of Gene Summary, in *proceedings of the BioNLP 2009 Workshop,* Boulder, USA 2009

6. Neveol A, Kim W, Wilbur WJ, **Lu Z***: Exploring two biomedical text genres for disease recognition, in *proceedings of the BioNLP 2009 Workshop,* Boulder, USA 2009

7. Chen S, Huang M, **Lu Z**: Summarizing Documents by Measuring the Importance of a Subset of Vertices within a Graph, in *proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intel ligence and Intel ligent Agent Technology*, Italy, 2009

8. **Lu Z*** and Wilbur WJ: Improving accuracy for identifying related PubMed queries by an integrated approach, *Journal of Biomedical Informatics*, 2008.

9. **Lu Z***, Kim W, Wilbur WJ: Evaluation of Query Expansion Using MeSH in PubMed, *Information Retrievel*, Volume 12, Issue 1, pages 69-80 2009.

10. **Lu Z***, Kim W, Wilbur WJ: Evaluating Relevance Ranking Strategies for MEDLINE Retrieval, *JAMIA*, 2009. *(recognized as one of the 9 outstanding papers from a total of 457 papers submitted to the AMIA 2008 Annual Symposium, and invited for publication in JAMIA)*.

11. Morgan AA, **Lu Z**, Wang X, Cohen AM, Fluck J, Ruch P, Divoli A, Fundel K, Leaman R, Hakenberg J, Sun C, Liu H, Torres R, Krauthammer M, Lau WW, Liu H, Hsu C, Schuemie M, Cohen KB, Hirschman L: Overview of BioCreative II Gene Normalization, *Genome Biology*, 2008, 9(Suppl 2):S2

12. Baumgartner WA, Jr, **Lu Z**, Johnson HL, Caporaso JG, Paquette J, Lindemann A, White EK, Medvedeva O, Cohen KB, Hunter L: Concept recognition for extracting protein interaction relations from biomedical text, *Genome Biology,* 2008, 9(Suppl 2):S8

13. Bethard S, **Lu Z***, Martin JH, Hunter L: Semantic Role Labeling for Protein Transport Predicates, *BMC Bioinformatics*, 9:277, 2008

14. Hunter L, **Lu Z**, Firby J, Baumgartner WA, Jr., Johnson HL, Ogren PV, Cohen KB: OpenDMAP: An open-source ontology-driven concept analysis engine, with applications to capturing knowledge regarding protein transport, protein interaction, and cell-specific gene expression, *BMC Bioinformatics*, 9(1):78, 2008

15. Baumgartner W, **Lu Z**, Johnson HL, Caporaso JG, Paquette J, Lindemann A, White EK, Medvedeva O, Cohen KB, and Hunter L: An integrated approach to concept recog-

nition in biomedical text, In *Proceedings of the Second BioCreative Challenge Evaluation Workshop*, 2007

16. **Lu Z***, Cohen KB, and Hunter L: GeneRIF Quality Assurance as Summary Revision, in *Proceedings of the Pacific Symposium on Biocomputing* 12:269-280, 2007

17. Caporaso JG, Baumgartner W, Kim H, **Lu Z**, Johnson H, Medvedeva O, Lindemann A, Fox L, White EK, Cohen KB, Hunter L: Concept Recognition, Information Retrieval, and Machine Learning in Genomics Question-Answering, in *Proceedings of The Fifteenth Text Rtrieval Conference*, (TREC 2006)

18. **Lu Z**, Bada M, Ogren P, Cohen KB, and Hunter L: Improving Biomedical Corpus Annotation Guidelines, in *Proceedings of the Joint BioLink and 9th Bio-Ontologies Meeting*, 5 August 2006; Fortaleza, Brazil

19. **Lu Z***, Cohen KB, and Hunter L: Finding GeneRIFs via Gene Ontology Terms, in *Proceedings of the Pacific Symposium on Biocomputing*, 3-7 January 2006, Maui, Hawaii. 2006:52-63

20. Johnson HL, Cohen KB, Baumgartner WA, **Lu Z**, Bada M, Kester T, Kim H, and Hunter L: Evaluation of Lexical Methods for Detecting Relationships Between Concepts From Multiple Ontologies, in *Proceedings of the Pacific Symposium on Biocomputing*, 3-7 January 2006, Maui, Hawaii. 2006:28-49

21. **Lu, Z*** and Hunter, L: GO Molecular Function Terms are Predictive of Subcellular Localization, in *Proceedings of the Pacific Symposium on Biocomputing*, 4-8 January 2005, Big Island of Hawaii. 2005:151-161

22. Szafron D, Lu P, Greiner R, Wishart DS, Poulin B, Eisner R, **Lu Z**, Anvik J, Macdonell C, Fyshe A, Meeuwis D: Proteome Analyst: custom predictions with explanations in a web-based tool for high-throughput proteome annotations. *Nucleic Acids Research* 2004 Jul 1, 32:W365-71

23. **Lu Z**, Szafron D, Greiner R, Lu P, Wishart DS, Poulin B, Anvik J, Macdonell C, Eisner R: Predicting subcellular localization of proteins using machine-learned classifiers. *Bioinformatics.* 2004 Mar 1; 20(4): 547-56

24. Szafron D, Greiner R, Lu P, Wishart DS, **Lu Z**, Poulin B, Eisner R, Anvik J, Macdonell C: Proteome Analyst Transparent High-throughput Protein Function Prediction. In *Proceedings of the International Conference on Machine Learning, Workshop on Machine Learning in Bioinformatics*, August 2003, Washington DC, USA

25. Sander J, Qin X, **Lu Z**, Niu N and Kovarsky A: Automatic Extraction of Clusters from Hierarchical Clustering Representations, in *Proceedings of The 7th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, April 2003, Seoul Korea.

PUBLISHED
REFEREED
ABSTRACTS

Mork JG, Bodenreider O, Demner-Fushman D, Dogan RI, Lang FM, **Lu Z**, Neveol A, Peters L, Shooshan SE, Aronson AR, NLM's I2B2 Tool System Description, *the Third i2b2 Workshop on Natural Language Processing Challenges for Clinical Records*, November, San Francisco, 2009.

Murray GC, Lin J, Wilbur WJ, **Lu Z**, Users' Adjustments to Unsuccessful Queries in Biomedical Search, *the 9th Joint Conference on Digital Library*, Austin, Texas, 2009.

Lu Z, Kim W, Wilbur WJ, Evaluating Relevance Ranking and Query Expansion for MEDLINE Retrieval, *the 6th Annual Rocky Mountain Bioinformatics Conference*, Colorado, 2008.

Lu Z, Baumgartner WA, Caporaso JG, Cohen KB, Hunter L. Finding infomrative sentences in full-text journal articles, *the 4th Annual Rocky Mountain Bioinformatics Conference*, Colorado, 2006.

Caporaso JG, Baumgartner WA, Kim H, **Lu Z**, Johnson HL, Medvedeva O, Lindemann A, Fox L, White EK, Cohen KB, Hunter L. Information retrieval, question-answering, machine learning, and concept recognition in TREC Genomics 2006, *the 4th Annual Rocky Mountain Bioinformatics Conference*, Colorado, 2006.

Lu Z, Baumgartner W, Bada M, Firby J, Ogren P, Cohen K, Hunter L. Frame-based Analysis of Complex Relationships in Biomedical Text: Annotating and Extracting Protein Transport Information in GeneRIFs, *the 14th International Conference on Intelligent Systems for Molecular Biology*, Fortaleza, Brazil, 2006

Lu Z, Hunter L. Mining Protein Transport Data from GeneRIFs, *the 3rd Annual Rocky Mountain Bioinformatics Conference*, Snowmass, Colorado, 2005

Lu Z, Ogren P, Dolbey A, Hunter L. PTKB: Protein Transport Knowledge Base, *the 13th International Conference on Intelligent Systems for Molecular Biology*, Detroit, Michigan, 2005

Lu Z, Hunter L. GO Molecular Function Terms are Predicative of Subcellular Localization, *the 2nd Annual Rocky Mountain Bioinformatics Conference*, Snowmass, Colorado, 2005

Lu Z, Using Proteome Analyst for Protein Subcellular Localization Predication, *the 1st Annual Rocky Mountain Bioinformatics Conference*, Snowmass, Colorado, 2004

Lu Z, Szafron D, Lu P, Wishart DS, Poulin B, Anvik J, Macdonell C, Eisner R. Using Proteome Analyst for Protein Subcellular Localization Predication, *the 3rd International Conference of the Canadian Proteomics Initiative*, Vancouver, BC, 2003

Lu Z, Wu X, Greiner R. Protein Sequence Classification in Proteome Analyst, *the 10th International Conference on Intelligent System for Molecular Biology*, Edmonton, Alberta, 2002

Neveol A, Islamaj-Dogan R, **Lu Z***, Quality, efficiency, satisfaction: a study of semi-automatic semantic annotation of PubMed queries, *submitted*